

# Coherent-to-Diffuse Power Ratio Estimation for Dereverberation

Andreas Schwarz\*, Walter Kellermann, *Fellow, IEEE*

**Abstract**—The estimation of the time- and frequency-dependent coherent-to-diffuse power ratio (CDR) from the measured spatial coherence between two omnidirectional microphones is investigated. Known CDR estimators are formulated in a common framework, illustrated using a geometric interpretation in the complex plane, and investigated with respect to bias and robustness towards model errors. Several novel unbiased CDR estimators are proposed, and it is shown that knowledge of either the direction of arrival (DOA) of the target source or the coherence of the noise field is sufficient for unbiased CDR estimation. The validity of the model for the application of CDR estimates to dereverberation is investigated using measured and simulated impulse responses. A CDR-based dereverberation system is presented and evaluated using signal-based quality measures as well as automatic speech recognition accuracy. The results show that the proposed unbiased estimators have a practical advantage over existing estimators, and that the proposed DOA-independent estimator can be used for effective blind dereverberation.

**Index Terms**—Spatial Coherence, Diffuse Noise Suppression, Diffuseness, Dereverberation, Reverberation Suppression

## I. INTRODUCTION

IT has been observed as early as 1969 that the measured spatial coherence between two microphones allows the discrimination between direct sound and reverberation [1]. A first signal enhancement algorithm based on this observation was proposed by Allen et al. in 1977 [2], where the magnitude of the coherence is estimated in the Short-Time Fourier Transform (STFT) domain and used as a gain for reverberation suppression. Other heuristic methods for noise reduction and dereverberation using coherence estimates have since been proposed [3]–[7]. Related methods have also been investigated for noise suppression in connection with beamforming, and postfilters which are statistically optimal under certain conditions have been proposed for the suppression of uncorrelated [8] and diffuse [9] noise.

More recently, explicit estimators for the ratio between direct and diffuse signal components, termed the coherent-to-diffuse power ratio (CDR), from short-time coherence estimates have been formulated [10], [11], based on the same assumptions as the earlier optimum postfilter derivations [9]. Also, results have since been generalized from omnidirectional microphones to other microphone directivities [12], [13] and spherical microphone arrays [14]. While these estimates can be

used for the formulation of postfilters for signal enhancement [15], which is the main application considered in this contribution, short-time CDR estimates (or the equivalent “diffuseness” measure) also have applications in parametric coding of spatial audio signals [16] and the extraction of spatial features for automatic speech recognition (ASR) [17].

In this contribution, the estimation of the CDR from the measured coherence between two omnidirectional microphones, and the application of the CDR estimates to dereverberation, is investigated. First, the signal model for the recording of a noisy or reverberant signal with two omnidirectional microphones is described, the relationship between signal and noise coherence models and the coherence of the mixed signal is given, and coherence models for the application to dereverberation are discussed. Then, several known CDR estimators are formulated in a common framework, illustrated using a geometric interpretation in the complex plane, and improved unbiased estimators are proposed. It is shown that knowledge of either the target signal direction or the noise coherence is sufficient for an unbiased CDR estimation, and estimators are proposed for the cases of unknown target signal direction and unknown noise coherence. Finally, the CDR estimators are applied in a postfilter for reverberation suppression and evaluated by processing reverberant speech and comparing ASR recognition accuracy as well as various signal quality measures. This paper builds on results published in a recent conference paper by the same authors, in which the novel estimators were initially proposed [15].

## II. SIGNAL MODEL

We consider the recording of a reverberant or noisy speech signal by two omnidirectional microphones with a spacing  $d$ , located in the same horizontal plane. The signal  $x_i(t)$  of the  $i$ -th microphone is composed of a desired signal component  $s_i(t)$  and an undesired component  $n_i(t)$  consisting of noise and/or late reverberation, i.e.,

$$x_i(t) = s_i(t) + n_i(t), \quad i = 1, 2. \quad (1)$$

The microphone, desired and noise signals are represented in the time-frequency (STFT) domain by the corresponding uppercase letters, i.e.,  $X_i(l, f)$ ,  $S_i(l, f)$  and  $N_i(l, f)$ , respectively, with the discrete-time frame index  $l$  and continuous frequency  $f$ , and are assumed to be short-time stationary. Using the representation in the STFT domain, the short-time auto- and cross-power spectra between two signals  $u(t)$  and  $v(t)$  are defined as

$$\Phi_{uv}(l, f) = \mathcal{E}\{U(l, f)V^*(l, f)\}, \quad (2)$$

A. Schwarz and W. Kellermann are with the Chair of Multimedia Communications and Signal Processing, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany (e-mail: schwarz@LNT.de; wk@LNT.de).

The authors would like to thank Opticom GmbH for providing PESQ evaluation software.

EDICS: AUD-SIRR, AUD-MAAE, AUD-SEN, AUD-ASAP

where  $\mathcal{E}$  is the expectation operator. It is assumed that the auto-power spectra of the signal components are the same at both microphones, i.e.,

$$\Phi_{s_1 s_1}(l, f) = \Phi_{s_2 s_2}(l, f) = \Phi_s(l, f), \quad (3)$$

$$\Phi_{n_1 n_1}(l, f) = \Phi_{n_2 n_2}(l, f) = \Phi_n(l, f). \quad (4)$$

Note that this assumption is generally appropriate for a plane wave as desired signal as well as for noise and late reverberation, but may in practice be impacted by the presence of early reflections causing destructive or constructive interference. The time- and frequency-dependent signal-to-noise ratio (SNR) of the microphone signals can be defined as

$$SNR(l, f) = \frac{\Phi_s(l, f)}{\Phi_n(l, f)}. \quad (5)$$

The complex spatial coherence functions of the desired signal and noise components are given by

$$\Gamma_s(f) = \frac{\Phi_{s_1 s_2}(l, f)}{\Phi_s(l, f)}, \Gamma_n(f) = \frac{\Phi_{n_1 n_2}(l, f)}{\Phi_n(l, f)}, \quad (6)$$

respectively, and are assumed to be time-invariant, i.e., dependent only on the spatial characteristics of the signal components. It is furthermore assumed that signal and noise are mutually orthogonal, such that

$$\Phi_x(l, f) = \Phi_s(l, f) + \Phi_n(l, f). \quad (7)$$

The complex spatial coherence of the mixed sound field can then be written as a function of the SNR and the signal and noise coherence functions:

$$\Gamma_x(l, f) = \frac{SNR(l, f)\Gamma_s(f) + \Gamma_n(f)}{SNR(l, f) + 1}. \quad (8)$$

This relationship is valid for any signal and noise coherence function. For the special case of a fully coherent desired signal component and diffuse noise, the term CDR or direct-to-diffuse ratio (DDR) is often used for the SNR. We will adopt the term CDR in the following. (8) can be rewritten as a parametric line equation in the complex plane, highlighting that  $\Gamma_x$  lies on a straight line connecting  $\Gamma_n$  and  $\Gamma_s$ :

$$\Gamma_x(l, f) = \Gamma_s(f) + \frac{1}{CDR(l, f) + 1}(\Gamma_n(f) - \Gamma_s(f)). \quad (9)$$

Note that the line parameter  $D(l, f) = [CDR(l, f) + 1]^{-1}$  is equivalent to the *diffuseness* defined in [18].

### III. COHERENCE MODELS FOR DEREVERBERATION

The desired and noise or reverberation components of the microphone signals are characterized by time-invariant coherence functions  $\Gamma_s(f)$  and  $\Gamma_n(f)$ , respectively. In the following, suitable models for these spatial coherence functions are discussed for the application to dereverberation.

#### A. Desired Signal

The desired signal component is modeled as a plane wave with the direction of arrival (DOA)  $\theta$  with respect to the microphone axis, where  $\theta = 0^\circ$  corresponds to broadside direction. The corresponding time-invariant coherence function is given by

$$\Gamma_s(f) = \frac{\Phi_{s_1 s_2}(l, f)}{\Phi_s(l, f)} = e^{jkd \sin(\theta)} = e^{j2\pi f \Delta t}, \quad (10)$$

with the time difference of arrival (TDOA)  $\Delta t = d \sin(\theta)/c$ , the wavenumber  $k = 2\pi f/c$  and the speed of sound  $c$ . This coherence function always has a magnitude of one, and is equal to one for  $\Delta t = 0$ .

#### B. Reverberation as Isotropic Sound Field

In array signal processing, environmental noise is often modeled by the superposition of an infinite number of uncorrelated, spatially distributed noise sources. In applications like underwater acoustics or radio communication, this model is motivated by the presence of many independent noise and interfering sources around the receiver [19]. The most common assumption for the spatial distribution is a sphere centered around the receiver, which corresponds to what is known as a *diffuse* or *spherically isotropic* noise field. The spatial coherence function between two omnidirectional sensors in a diffuse noise field is real-valued and given by

$$\Gamma_{\text{diffuse}}(f) = \frac{\sin(kd)}{kd} = \frac{\sin(2\pi f d/c)}{2\pi f d/c}. \quad (11)$$

While diffusivity of the noise field is easily motivated in the aforementioned scenarios, a few more considerations are necessary for the modeling of a reverberation component originating from a single excitation signal. Since acoustic transmission within a room is generally assumed to be linear and time-invariant, a reverberant signal can be modeled by the convolution of a source signal with a time-invariant room impulse response (RIR) [20]. The reverberant signals recorded at two points in space, i.e., by two microphones, are therefore linearly related, and the theoretical coherence function between these two signals is equal to one. However, when limited observation windows are considered, and the excitation signal has a limited temporal correlation, reflections with different delays can be approximated as uncorrelated sources. This uncorrelated scattering assumption is widely used in mobile radio communications [21] and underwater acoustics [22], and is useful in room acoustics as well, where it has been observed that the sound field in a reverberant room appears as an approximately diffuse sound field [23], [24]. The plausibility of the diffuseness assumption for reverberation can be visualized using the image source model [25]: for higher reflection orders, the angular distribution of the image sources becomes increasingly isotropic. Furthermore, given a limited observation window length, the delayed reflected versions of the source signal are increasingly decorrelated with increasing reflection orders. Based on this idea, we can predict a number of factors which contribute to how well the model of diffuseness is fulfilled: a large room contributes

to the uncorrelatedness of the image sources, due to larger relative delays between reflections; highly reflective surfaces contribute to the presence of many image sources with similar power, since the power contributed by reflections decays more slowly with the reflection order; and low temporal correlation of the source signal contributes to low correlation between the delayed reflections. Some of these effects are illustrated in Section VI-B using measured and simulated RIRs.

In real rooms, effects like diffraction, diffuse reflection [20], and potentially time-variant effects [26] may further contribute to the randomization of delays and incidence angles of reflections and therefore increase the diffuseness of the reverberation sound field. However, as shown later, the image source model is sufficient to explain a wide range of practical effects which affect the reverberation coherence.

While the diffuse sound field model is the most common in room acoustics and signal enhancement, it has been observed that reverberant noise in rooms with highly absorbing floors and ceilings can be modeled more accurately by noise sources distributed in the horizontal plane, i.e., by a 2D isotropic (cylindrically isotropic) noise field, as opposed to a diffuse (spherically isotropic) noise field [27]. This noise field model consists of uncorrelated noise sources located on a circle around and in the same plane as the microphones (typically the horizontal plane), and is motivated by the rapid decay of all vertically propagating sound components due to the strong absorption at the floor and/or ceiling. The corresponding spatial coherence function for two omnidirectional microphones located in the same plane as the noise sources is the zeroth-order Bessel function of the first kind [23], [28]:

$$\Gamma_{2D\text{-iso}}(f) = J_0(kd) = J_0(2\pi fd/c). \quad (12)$$

Note that, both in the case of diffuse and 2D-isotropic noise fields, the coherence function is real-valued, since the spatial distribution of the sources is symmetric with respect to the microphone array axis.

In Section VI-B, the effects of room geometry and surface reflectivity on the coherence of the reverberation component are evaluated using RIRs generated with the image source method, and RIRs that were measured in different rooms.

#### IV. COHERENT-TO-DIFFUSE POWER RATIO ESTIMATION

For most proposed postfilters, the gain function has been formulated directly as a function of auto- and cross-power spectral estimates [8], [9], which are typically obtained from the microphone signals by recursive averaging:

$$\hat{\Phi}_{x_i x_j}(l, f) = \lambda \hat{\Phi}_{x_i x_j}(l-1, f) + (1-\lambda) X_i(l, f) X_j^*(l, f), \quad (13)$$

where  $\lambda$  is a constant between 0 and 1. We follow a different approach where we first derive an SNR estimate, which can then be used to apply any suppression technique such as the Wiener filter or spectral subtraction [29]. Furthermore, we write the estimate not as a function of auto- and cross-power spectral estimates, but as a function of the estimated short-time spatial coherence, which allows additional insight into

the behavior of the estimator. The short-time coherence is estimated by

$$\hat{\Gamma}_x(l, f) = \frac{\hat{\Phi}_{x_1 x_2}(l, f)}{\sqrt{\hat{\Phi}_{x_1 x_1}(l, f) \hat{\Phi}_{x_2 x_2}(l, f)}}. \quad (14)$$

Since the focus is on estimating the SNR for a mixture of a fully coherent signal with  $|\Gamma_s(f)| = 1$  and isotropic noise with  $\Gamma_n \in \mathbb{R}$ , where typically  $\Gamma_n(f) = \Gamma_{\text{diffuse}}(f)$ , we use the term CDR instead of SNR for the quantity to be estimated in the following. For the application to dereverberation, the CDR is equivalent to the direct-to-reverberation power ratio (DRR), under the assumption that reverberant sound can be modeled as a mixture of a direct component and a perfectly diffuse reverberation component which are mutually uncorrelated, thus neglecting early reflections.

The aim is now to estimate the CDR from an estimate of the short-time spatial coherence  $\hat{\Gamma}_x(l, f)$ , exploiting the known coherence functions of the signal and/or noise component, and the relationship of these coherence models and the mixed sound field coherence to the CDR given by (9). Solving (9) for the CDR yields (for brevity, the time- and frequency-dependency is omitted in the following)

$$CDR = \frac{\Gamma_n - \Gamma_x}{\Gamma_x - \Gamma_s}, \quad (15)$$

or, reformulated as the diffuseness  $D$ ,

$$D = \frac{1}{CDR + 1} = \frac{\Gamma_x - \Gamma_s}{\Gamma_n - \Gamma_s}. \quad (16)$$

Although  $\Gamma_x$  and  $\Gamma_s$  may be complex, the CDR and diffuseness are real-valued quantities; however, when inserting a coherence estimate  $\hat{\Gamma}_x$  for  $\Gamma_x$  in (15), the resulting values are in general complex-valued, due to mismatch between the coherence models and the actual acoustic conditions, and the variance of the coherence estimate. Estimating the CDR by direct application of (15) is therefore not feasible, which is why a number of different estimator implementations, which yield a positive, real-valued CDR estimate for all possible values of  $\hat{\Gamma}_x$ ,  $|\hat{\Gamma}_x| \leq 1$ , have been proposed.

In the following, first, the interpretation of the estimator behavior in the complex plane is discussed. Then, existing and novel approaches to CDR estimation are analyzed. For an easier comparison, the estimators are reformulated as a function of only the coherence estimate  $\hat{\Gamma}_x$  and the assumed coherence models  $\hat{\Gamma}_s$  and  $\hat{\Gamma}_n$ , where  $\hat{\Gamma}_s$  is the direct signal coherence computed according to (10) from an a-priori known or estimated TDOA  $\hat{\Delta}t$ , and  $\hat{\Gamma}_n$  is assumed to match the diffuse coherence model (11). We start with methods which make use of both  $\hat{\Gamma}_s$  and  $\hat{\Gamma}_n$ , i.e., exploit information on the DOA and the noise coherence, continue with DOA-independent estimators which exploit only the knowledge of  $\hat{\Gamma}_n$ , and finally propose a CDR estimator for the case of available signal coherence  $\hat{\Gamma}_s$ , but unknown noise coherence. Table I summarizes the presented estimators and their main properties. Finally, estimator bias and robustness are evaluated.

### A. Interpretation of Estimator Behavior in the Complex Plane

Fig. 1 shows the output of the estimators which are described in the following sections in the complex plane of possible coherence values  $\hat{\Gamma}_x$ . Results for a direct signal TDOA  $\Delta t = 0$  (broadside) are shown in the first row, while in the second row, results are shown for  $\Delta t = \frac{1}{5f}$ . For all estimators,  $\tilde{\Gamma}_s = \Gamma_s$ ,  $\tilde{\Gamma}_n = \Gamma_n$  is assumed. The symbol  $\circ$  marks the coherence of a fully coherent signal with the respective TDOA according to (10), while the symbol  $\times$  marks the coherence of an ideal diffuse signal given by (11). The straight white line between these points marks the theoretical coherence values which would occur under ideal conditions for different CDR values, according to (9). The *bias* of a CDR estimator is henceforth defined as the deviation of the estimator from (15) for coherence values along this line; i.e., an *unbiased* estimator should exactly match (15) for these values. This can be verified by inserting  $\Gamma_x$  according to (9) for  $\hat{\Gamma}_x$  into the estimator equation, which yields  $\widehat{CDR} = CDR$  for an unbiased estimator. Furthermore, since the coherence estimates  $\hat{\Gamma}_x$ , which are observed in practice, will not lie exactly on the line, a good estimator should also be *robust* in the sense that some deviations of the coherence estimate from the assumed model, e.g., caused by an imperfect DOA estimate, do not lead to large deviations of the CDR estimate. In Fig. 1, robustness can be seen in the change of the CDR estimate for coherence values slightly deviating from the line; if these changes are abrupt, as in Fig. 1b for coherence values close to the unit circle, this indicates non-robust behavior. While we do not derive a measure for the overall robustness of an estimator, which would require establishing a statistical model for the errors, we evaluate the behavior of the different estimators with coherence model errors in Section IV-E.

### B. CDR Estimation for Known DOA and Noise Coherence

Using the same model as described in Section II, McCowan and Boulard [9] derived the Wiener postfilter for a coherent signal in diffuse noise. Jeub et al. [30] evaluated this postfilter for the suppression of reverberation, and formulated a CDR

Table I  
OVERVIEW OF INVESTIGATED CDR ESTIMATORS, REQUIRED PRIOR INFORMATION (NOISE AND/OR SIGNAL COHERENCE) AND UNBIASEDNESS.

Estimator	Definition	Required	Unbiased
Jeub	$\frac{\tilde{\Gamma}_n - \text{Re}\{\tilde{\Gamma}_s^* \hat{\Gamma}_x\}}{\text{Re}\{\tilde{\Gamma}_s^* \hat{\Gamma}_x\} - 1}$	$\tilde{\Gamma}_n, \tilde{\Gamma}_s$	no
Thiergart 1	$\text{Re}\left\{\frac{\tilde{\Gamma}_n - \hat{\Gamma}_x}{\tilde{\Gamma}_x - \tilde{\Gamma}_s}\right\}$	$\tilde{\Gamma}_n, \tilde{\Gamma}_s$	yes
Proposed 1	$\frac{\text{Re}\{\tilde{\Gamma}_s^* (\tilde{\Gamma}_n - \hat{\Gamma}_x)\}}{\text{Re}\{\tilde{\Gamma}_s^* \hat{\Gamma}_x\} - 1}$	$\tilde{\Gamma}_n, \tilde{\Gamma}_s$	yes
Proposed 2	$\frac{1 - \tilde{\Gamma}_n \cos(\arg(\tilde{\Gamma}_s))}{ \tilde{\Gamma}_n - \tilde{\Gamma}_s } \left  \frac{\tilde{\Gamma}_s^* (\tilde{\Gamma}_n - \hat{\Gamma}_x)}{\text{Re}\{\tilde{\Gamma}_s^* \hat{\Gamma}_x\} - 1} \right $	$\tilde{\Gamma}_n, \tilde{\Gamma}_s$	yes
Thiergart 2	$\text{Re}\left\{\frac{\tilde{\Gamma}_n - \hat{\Gamma}_x}{\tilde{\Gamma}_x - e^{j \arg \tilde{\Gamma}_x}}\right\}$	$\tilde{\Gamma}_n$	no
Proposed 3	(25)	$\tilde{\Gamma}_n$	yes
Proposed 4	(27)	$\tilde{\Gamma}_s$	yes

estimate based on the same model [10]. Both McCowan and Jeub rely on the assumption that the direct signal is time-aligned in both microphones, which can be achieved by applying a delay corresponding to the TDOA estimate  $\hat{\Delta t}$  to one of the channels [30]. In the STFT domain, this delay is equivalent to a phase rotation of the cross-power spectrum (assuming that the delay is significantly shorter than the transform length), and can therefore be represented in the CDR estimator equation by multiplying the complex rotation factor  $e^{-j2\pi f \hat{\Delta t}} = \tilde{\Gamma}_s^*$  with the coherence estimate  $\hat{\Gamma}_x$ . This allows the formulation of the CDR estimator including time alignment as a function of only  $\hat{\Gamma}_x$ ,  $\tilde{\Gamma}_s$  and  $\tilde{\Gamma}_n$ :

$$\begin{aligned} \widehat{CDR}_{\text{Jeub}}(l, f) &= \max \left( 0, \frac{\tilde{\Gamma}_n - \text{Re}\{e^{-j2\pi f \hat{\Delta t}} \hat{\Gamma}_x\}}{\text{Re}\{e^{-j2\pi f \hat{\Delta t}} \hat{\Gamma}_x\} - 1} \right) \\ &= \max \left( 0, \frac{\tilde{\Gamma}_n - \text{Re}\{\tilde{\Gamma}_s^* \hat{\Gamma}_x\}}{\text{Re}\{\tilde{\Gamma}_s^* \hat{\Gamma}_x\} - 1} \right). \end{aligned} \quad (17)$$

The maximum operation is required to prevent negative results for the CDR estimate. This estimator is unbiased for  $\tilde{\Gamma}_s = 1$ , i.e.,  $\hat{\Delta t} = 0$ . However, for non-zero TDOAs, the phase rotation of the coherence estimate  $\hat{\Gamma}_x$  does not only affect the direct signal component, but also the coherence of the diffuse signal

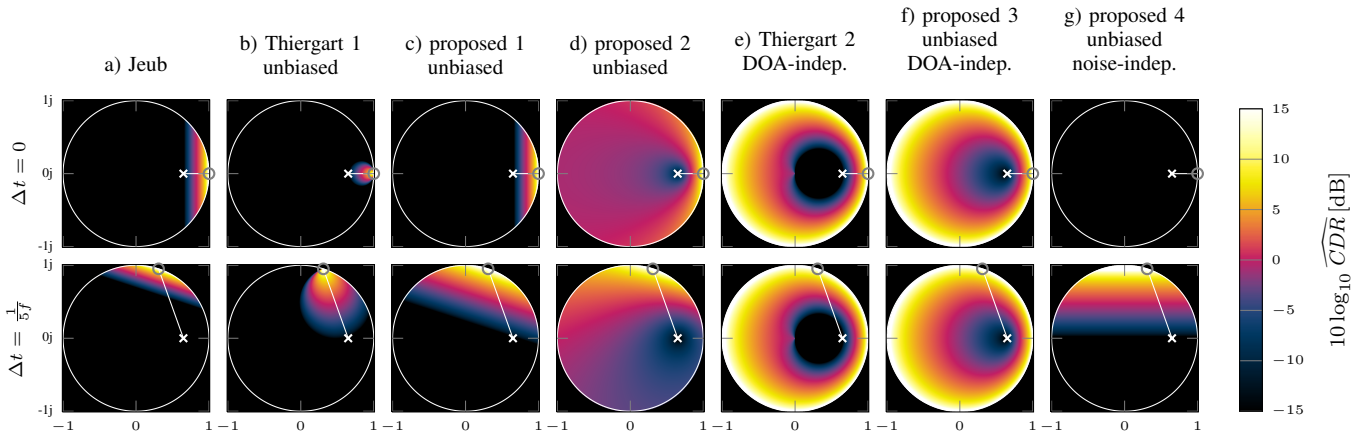


Figure 1. Coherent-to-diffuse power ratio estimates obtained from different estimators (columns) as a function of the complex spatial coherence estimate  $\hat{\Gamma}_x$ . The theoretical coherence of fully coherent ( $\Gamma_s$ ) and fully diffuse ( $\Gamma_n$ ) signals is marked by  $\circ$  and  $\times$ , respectively, while the theoretical coherence of mixed signals lies on the connecting line. Estimators are computed using  $\tilde{\Gamma}_s = \Gamma_s$ ,  $\tilde{\Gamma}_n = \Gamma_n$ . Parameters  $d = 8$  cm,  $f = 1$  kHz, different TDOAs (rows).

component. Since this is not accounted for by this estimator, the estimate is biased for non-zero TDOAs. The estimator is illustrated in Fig. 1a.

Thiergart et al. [11], [13] proposed to estimate the CDR by directly inserting the target signal coherence estimate  $\tilde{\Gamma}_s$  into (15), and taking the real part:

$$\widehat{CDR}_{\text{Thiergart1}}(l, f) = \max \left( 0, \text{Re} \left\{ \frac{\tilde{\Gamma}_n - \hat{\Gamma}_x}{\hat{\Gamma}_x - \tilde{\Gamma}_s} \right\} \right). \quad (18)$$

While this estimator is unbiased, it was found to be very sensitive towards phase deviations of the coherence estimate from the ideal model [13]. For a measured coherence with a magnitude close to one, even a small phase difference between  $\hat{\Gamma}_x$  and  $\Gamma_s$  can have a large effect on the CDR estimate. This can be seen in Fig. 1b, where, unlike in Fig. 1a, the CDR for coherence values close to the unit circle sharply drops to zero, and is shown in more detail later.

Based on (17), an unbiased CDR estimator can be formulated [15]. The diffuse coherence model is first corrected to account for the phase rotation of the coherence estimate by multiplying the diffuse noise coherence  $\tilde{\Gamma}_n$  with the phase term  $e^{-j2\pi f \hat{\Delta}t}$  as well, which removes the bias of the estimator, while preserving the robust properties of (17) against phase errors (see Fig. 1c):

$$\begin{aligned} \widehat{CDR}_{\text{prop1}}(l, f) &= \max \left( 0, \frac{\text{Re}\{e^{-j2\pi f \hat{\Delta}t} \tilde{\Gamma}_n - e^{-j2\pi f \hat{\Delta}t} \hat{\Gamma}_x\}}{\text{Re}\{e^{-j2\pi f \hat{\Delta}t} \hat{\Gamma}_x\} - 1} \right) \\ &= \max \left( 0, \frac{\text{Re}\{\tilde{\Gamma}_s^* (\tilde{\Gamma}_n - \hat{\Gamma}_x)\}}{\text{Re}\{\tilde{\Gamma}_s^* \hat{\Gamma}_x\} - 1} \right). \end{aligned} \quad (19)$$

This estimator is identical to (17) for  $\tilde{\Gamma}_s = 1$ , i.e.,  $\hat{\Delta}t = 0$ . Note that an equivalent CDR estimate can be derived from the maximum likelihood noise variance estimator which was proposed in [31] and applied to noise reduction in [32].

For a second, heuristically motivated variant of an unbiased estimator, the real part in the numerator of (19) and the max operator are first replaced by the magnitude of the entire term. The resulting estimator was found to lead to an increased performance for the application to dereverberation [33]:

$$\widehat{CDR}'_{\text{prop2}}(l, f) = \left| \frac{\tilde{\Gamma}_s^* (\tilde{\Gamma}_n - \hat{\Gamma}_x)}{\text{Re}\{\tilde{\Gamma}_s^* \hat{\Gamma}_x\} - 1} \right|. \quad (20)$$

This estimator however has a small bias for non-zero TDOAs; a correction term for this bias can be computed by inserting (9) into (20) and solving for  $\frac{CDR'}{CDR'_{\text{prop2}}}$ . The bias-compensated estimator is then given by

$$\widehat{CDR}_{\text{prop2}}(l, f) = \frac{1 - \tilde{\Gamma}_n \cos(\arg(\tilde{\Gamma}_s))}{|\tilde{\Gamma}_n - \tilde{\Gamma}_s|} \widehat{CDR}'_{\text{prop2}}(l, f), \quad (21)$$

and is illustrated in Fig. 1d. Compensation of this small bias however only has a negligible effect on practical performance.

The derivation of these estimators shows that, when both knowledge of the signal and noise coherence are available, several different unbiased CDR estimators can be implemented. The reason for this is that the requirement of unbiasedness only defines the behavior of the estimator for coherence values matching the model given by (9), i.e., the values on the line in Fig. 1, while allowing arbitrary behavior for other coherence values. While the second proposed unbiased variant has significant practical advantages, as shown in the qualitative analysis of the estimator behavior in Section IV-E and the signal-based evaluation in Section VI, it does not seem to be optimal in any sense. A possible direction for future work would therefore be to establish a statistical model for the deviations of  $\hat{\Gamma}_x$  from the theoretical model given by (9), and derive a correspondingly optimized unbiased estimator.

### C. CDR Estimation for Unknown DOA

The previously shown methods rely on prior knowledge or an estimate of the target DOA. As an alternative, Thiergart et al. [11], [13] proposed to use the instantaneous phase of the estimated cross-power spectrum  $\hat{\Phi}_{x_1 x_2}$  as a phase estimate for the direct signal model, i.e.,  $\tilde{\Gamma}_s = e^{j \arg \hat{\Phi}_{x_1 x_2}}$ , thus removing the need for explicit DOA estimation to obtain  $\tilde{\Gamma}_s$ . Since, according to (14),  $\arg \hat{\Gamma}_x = \arg \hat{\Phi}_{x_1 x_2}$ , this estimator can be formulated as a function of only the coherence estimate  $\hat{\Gamma}_x$  and the noise coherence  $\tilde{\Gamma}_n$ :

$$\widehat{CDR}_{\text{Thiergart2}}(l, f) = \max \left( 0, \text{Re} \left\{ \frac{\tilde{\Gamma}_n - \hat{\Gamma}_x}{\hat{\Gamma}_x - e^{j \arg \hat{\Gamma}_x}} \right\} \right). \quad (22)$$

However, the instantaneous phase of the mixture is not an unbiased estimate of the phase of the direct signal component, since, for low CDR values, the coherence of the mixture is dominated by the coherence of the diffuse signal component [13], which is real-valued, i.e., has a phase of zero. For  $\theta \neq 0^\circ$ , the estimator is therefore biased. The behavior of the estimator is illustrated in Fig. 1e.

As shown in [15], it is possible to derive an unbiased CDR estimator which does not require an estimate of the source DOA, since the knowledge that  $|\Gamma_s| = 1$ , i.e., that the direct signal is fully coherent, is sufficient to solve (15). This can be explained using a geometric interpretation: according to (9),  $\Gamma_x$ ,  $\Gamma_s$  and  $\Gamma_n$  all lie on a straight line in the complex plane, and it is furthermore known that  $\Gamma_s$  lies on the unit circle and  $\Gamma_n$  on the real axis.  $\Gamma_s$  can therefore be obtained by the intersection of the line through  $\Gamma_n$  and  $\Gamma_x$  with the unit circle, and inserted into (15). An alternative way of obtaining this solution is by solving (9) for  $\Gamma_s$  and setting the magnitude to 1:

$$|\Gamma_s| = |\Gamma_x - (\Gamma_n - \Gamma_x) CDR| \stackrel{!}{=} 1, \quad (23)$$

$$\widehat{CDR}_{\text{prop3}}(l, f) = \frac{\tilde{\Gamma}_n \text{Re}\{\hat{\Gamma}_x\} - |\hat{\Gamma}_x|^2 - \sqrt{\tilde{\Gamma}_n^2 \text{Re}\{\hat{\Gamma}_x\}^2 - \tilde{\Gamma}_n^2 |\hat{\Gamma}_x|^2 + \tilde{\Gamma}_n^2 - 2\tilde{\Gamma}_n \text{Re}\{\hat{\Gamma}_x\} + |\hat{\Gamma}_x|^2}}{|\hat{\Gamma}_x|^2 - 1} \quad (25)$$

which leads to a quadratic equation for the CDR:

$$(|\Gamma_x|^2 - 1)CDR^2 - 2\text{Re}\{\Gamma_x(\Gamma_n - \Gamma_x)^*\}CDR + |\Gamma_n - \Gamma_x|^2 = 0. \quad (24)$$

Taking the positive of both possible solutions yields the unbiased DOA-independent CDR estimator which is given by (25) and illustrated in Fig. 1f. In contrast to the DOA-dependent estimators, where an infinite number of unbiased estimators exists, the DOA-independent estimator is uniquely determined by the requirement of unbiasedness.

#### D. CDR Estimation for Unknown Noise Coherence

From the geometric interpretation of the coherence of mixed sound fields it can be analogously concluded that knowledge of  $\Gamma_n$  is not required when  $\Gamma_s$  is known, since the noise coherence is assumed to be real and therefore determined by the intersection of the real axis and the line through  $\Gamma_s$  and  $\Gamma_x$ . Using  $\text{Im}\{\Gamma_n\} = 0$ ,  $\Gamma_n$  can therefore be eliminated from (15), resulting in

$$CDR = \frac{\text{Im}\{\Gamma_x\}}{\text{Im}\{\Gamma_s\} - \text{Im}\{\Gamma_x\}}. \quad (26)$$

When using this formulation with the estimates  $\hat{\Gamma}_x$  and  $\tilde{\Gamma}_s$  as an estimator for the CDR, practical problems occur in cases where, due to model mismatch and coherence estimation errors, the imaginary part of the coherence estimate  $\text{Im}\{\hat{\Gamma}_x\}$  has either values with a larger magnitude than  $\text{Im}\{\tilde{\Gamma}_s\}$ , or a different sign, in which case this equation would not yield a meaningful result. For this reason, the CDR estimate is continuously extended into these two problematic regions by returning an infinite CDR in the former case, and a CDR of zero in the latter case. The final proposed estimator is then given by

$$\widehat{CDR}_{\text{prop4}}(l, f) = \begin{cases} \infty, & \text{for } \frac{\text{Im}\{\hat{\Gamma}_x\}}{\text{Im}\{\tilde{\Gamma}_s\}} \geq 1 \\ \frac{\text{Im}\{\hat{\Gamma}_x\}}{\text{Im}\{\tilde{\Gamma}_s\} - \text{Im}\{\hat{\Gamma}_x\}}, & \text{for } 0 < \frac{\text{Im}\{\hat{\Gamma}_x\}}{\text{Im}\{\tilde{\Gamma}_s\}} < 1 \\ 0, & \text{for } \frac{\text{Im}\{\hat{\Gamma}_x\}}{\text{Im}\{\tilde{\Gamma}_s\}} \leq 0. \end{cases} \quad (27)$$

An inherent constraint that limits practical applicability of this estimator is that  $\arg \Gamma_s \neq 0$ , since otherwise the imaginary parts disappear; i.e., the estimator is not usable for  $\Delta t = 0$ , and increasingly sensitive towards estimation errors for small TDOAs. The estimator is visualized in Fig. 1g. Note that in [34] a noise power spectrum estimate was derived in a similar way from the imaginary part of a cross-power spectrum.

#### E. Evaluation of Estimator Bias and Robustness

To illustrate the bias of the estimators  $\widehat{CDR}_{\text{Jeub}}$ , the uncompensated estimator  $\widehat{CDR}_{\text{prop2}}$  and  $\widehat{CDR}_{\text{Thiergart,2}}$ , Fig. 2 compares the true CDR value and the different estimates for mixtures of coherent and ideally diffuse signals for a TDOA  $\Delta t = \frac{1}{5f}$  (corresponding to the values along the white line in Fig. 1, second row). The proposed estimators are all unbiased, as is the DOA-dependent estimator proposed by Thiergart et al. (18). The estimator by Jeub et al. (17) and the DOA-independent estimator by Thiergart et al. (22) both have a

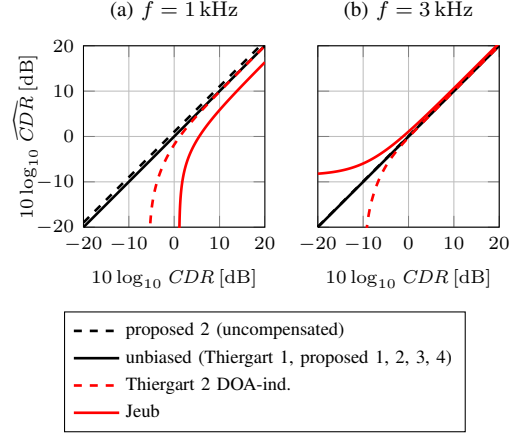


Figure 2. Comparison of true CDR and estimated CDR. Parameters  $d = 8$  cm,  $\Delta t = \frac{1}{5f}$ ,  $f = 1$  kHz (left), 3 kHz (right).

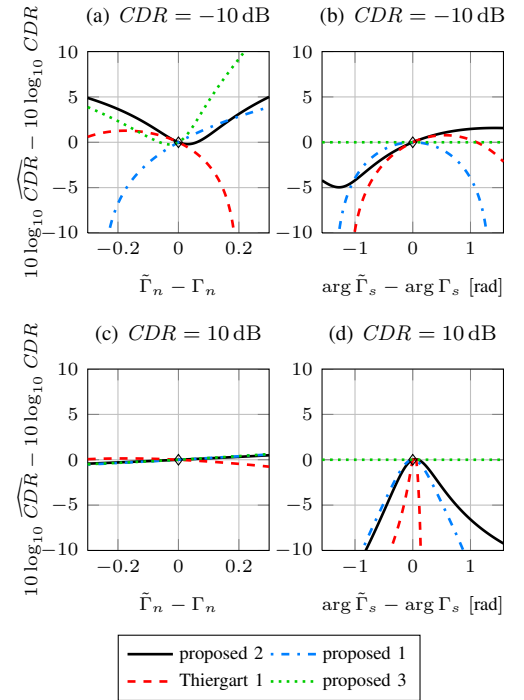


Figure 3. CDR estimation error for noise and direct signal coherence model errors. Parameters  $d = 8$  cm,  $\Delta t = \frac{1}{5f}$ ,  $f = 1$  kHz.

significant bias, with the former under- or overestimating the CDR depending on the values of  $\Delta t$  and  $f$ , and the latter always underestimating the CDR. Also shown is the uncompensated version of the proposed estimator 2 (20), which has a small, TDOA- and frequency-dependent bias (for  $f = 3$  kHz, the difference to the unbiased case is too small to be noticeable in the plot).

Fig. 3 shows the CDR estimation error for cases where the actual coherence of the noise  $\Gamma_n$  or the direct signal component  $\Gamma_s$  deviates from the assumed coherence models  $\tilde{\Gamma}_n$  and  $\tilde{\Gamma}_s$ , respectively. Fig. 3a and b show the error for a low CDR of  $-10$  dB, while c and d show results for a high CDR of  $10$  dB. The DOA-independent estimator  $\widehat{CDR}_{\text{prop3}}$  is naturally unaffected by the phase error of the direct signal

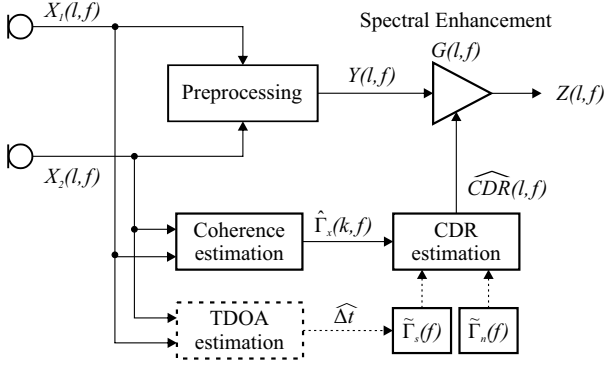


Figure 4. Coherence-based noise and reverberation suppression system consisting of a preprocessor and a CDR-based postfilter.

coherence model, as seen in Fig. 3b and d; however, for errors of the noise coherence, the CDR is quickly overestimated by the DOA-independent estimator (see Fig. 3a). The estimator  $\widehat{CDR}_{\text{Thiergart,1}}$  has the problem of reacting strongly to small phase deviations when the CDR is high (see Fig. 3d). Comparing the different unbiased DOA-dependent variants  $\widehat{CDR}_{\text{prop1}}$  and  $\widehat{CDR}_{\text{prop2}}$ , it can be stated that  $\widehat{CDR}_{\text{prop2}}$  seems slightly more tolerant towards model errors, which could explain the better performance of this estimator for signal enhancement.

## V. APPLICATION TO SPEECH ENHANCEMENT

Fig. 4 shows the structure of the proposed reverberation or diffuse noise suppression system based on short-time CDR estimates. First, the microphone signals are combined by averaging the squared magnitudes and using the phase from one of the microphone signals:

$$Y(l, f) = \frac{1}{2} \sqrt{|X_1(l, f)|^2 + |X_2(l, f)|^2} \cdot e^{j \arg X_1(l, f)}. \quad (28)$$

Spatial magnitude averaging in the STFT domain is typically used to reduce the variance of spectral estimates for the computation of microphone array postfilters [9], but has also been used as a preprocessor for signal enhancement [35]. It is used here with the purpose of reducing the variations in the transfer function which are caused by constructive and destructive interference of early reflection components with the direct path. For the computation of the coherence-based postfilter gain  $G(l, f)$ , short-time estimates  $\hat{\Gamma}_x(l, f)$  of the spatial coherence are first obtained according to (14) from spectra which have been estimated by recursive averaging. From the coherence, the CDR is estimated based on models for the direct signal and/or reverberation coherence, where the direct signal coherence is derived from a known or estimated TDOA, and the reverberation coherence is assumed to be known. A postfilter gain is then computed using spectral magnitude subtraction [29]:

$$G(l, f) = \max \left\{ G_{\min}, 1 - \sqrt{\frac{\mu}{\widehat{CDR}(l, f) + 1}} \right\}, \quad (29)$$

with the oversubtraction factor  $\mu$  and the gain floor  $G_{\min}$ . The output signal is computed by applying the postfilter gain to

the preprocessed signal  $Y(l, f)$ , i.e.,  $Z(l, f) = G(l, f)Y(l, f)$ , and transformed back into the time domain. Since the preprocessor does not have any spatial filtering effect, the postfilter gain can be directly applied to the preprocessor output, and does not require a correction to account for spatial filtering, as it would be the case for a beamformer as preprocessor [8].

Note that, when employing a DOA-independent CDR estimator, the proposed signal enhancement system is completely independent of the DOA of the target signal.

## VI. EVALUATION

In the following, the spatial properties of reverberation are first evaluated using simulated and measured RIRs, in order to verify the assumptions made in Sect. III-B. Then, the estimation accuracy of the CDR estimators and the effect of the proposed CDR-based dereverberation system are evaluated.

A MATLAB implementation of the proposed CDR estimators and signal enhancement scheme is provided online<sup>1</sup>.

### A. Setup and Parameters

For the main evaluation, sets of measured RIRs from three rooms are used:

- Room A: 6 m × 6 m × 3 m, partially closed curtains on walls,  $T_{60} \approx 0.4$  s
- Room B: 7 m × 11 m × 3 m (lecture hall),  $T_{60} \approx 1$  s
- Room C: 54 m × 7 m × 3 m (large foyer),  $T_{60} \approx 3.5$  s

The reverberation time  $T_{60}$  was measured from the energy decay curve of the RIR. In each room, RIRs were measured for 40-70 different source positions in  $l = 1, 2$  and 4 m distance from the microphones, in the angular range  $\theta = -90 \dots 90^\circ$ . Microphones are spaced  $d = 8$  cm apart.

Additionally, the RIRs that were used in the REVERB challenge [36] for the generation of multi-condition training data are evaluated. These RIRs were measured using an 8-channel circular microphone array with a diameter of 20 cm (corresponding to  $d = 8$  cm spacing between neighboring microphones) in 6 different rooms (SR1/2, MR1/2, LR1/2), for two source-microphone distances ( $\approx 0.5$  m and  $\approx 2$  m), and two different angles of the source w.r.t. the microphone array. The rooms have the following properties (note that SR2 and LR2 are the same rooms as A and B, respectively):

- SR1 (“Small Room 1”): variable reverberation room, 4.5 m × 3.5 m × 3 m,  $T_{60} \approx 0.2$  s
- SR2 (“Small Room 2”): room A, but curtains fully closed,  $T_{60} \approx 0.2$  s
- MR1 (“Medium Room 1”): same as SR1,  $T_{60} \approx 0.5$  s
- MR2 (“Medium Room 2”): meeting room, 5 m × 3.5 m × 3 m,  $T_{60} \approx 0.6$  s
- LR1 (“Large Room 1”): same as SR1,  $T_{60} \approx 0.8$  s
- LR2 (“Large Room 2”): room B

In the following, all processing takes place at a sampling rate of 16 kHz. For the transformation into the time-frequency domain and short-time spectral estimation, a DFT-based uniform filterbank with window length 1024, FFT size 512, and downsampling factor 128 is employed [37]. The short-time

<sup>1</sup><http://www.lms.lnt.de/files/publications/cdr-dereverb.zip>



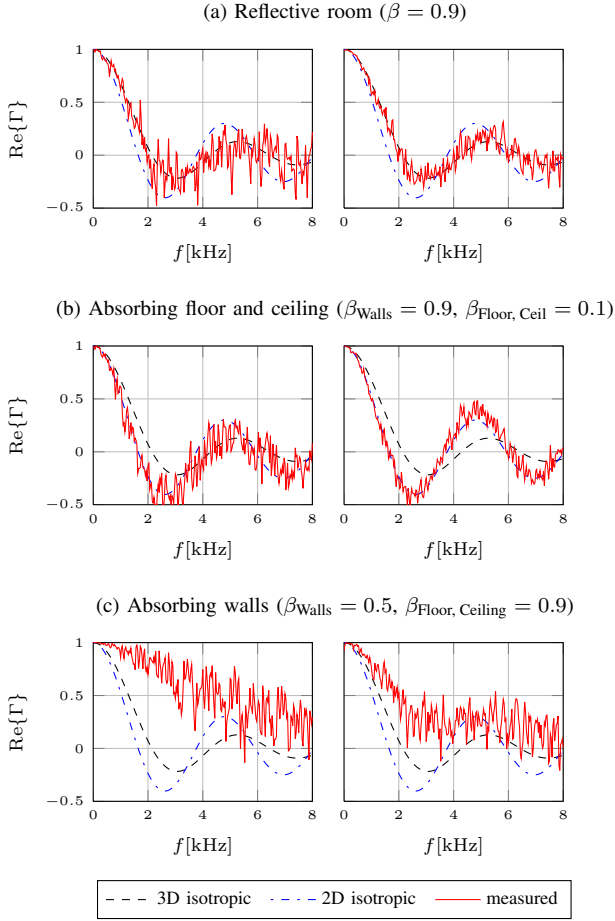


Figure 5. Spatial coherence estimated from the reverberation tail of simulated RIRs, averaged over 7 microphone pairs with spacing  $d = 8$  cm, for different reflection coefficients  $\beta$ , compared to coherence of diffuse and 2D isotropic sound fields. Left: small room ( $4 \times 3 \times 2.5$  m), right: large room ( $15 \times 18 \times 10$  m).

coherence estimates are obtained by recursive averaging of the auto- and cross-power spectra according to (13), with the forgetting factor  $\lambda = 0.68$ .

### B. Spatial Properties of Reverberation in Simulated and Measured Rooms

For the evaluation of the spatial characteristics of reverberation, we use simulated and measured RIRs. The reverberation tail of the RIRs is extracted by removing the initial part containing the direct path and early reflections (see Appendix), using a typical value of  $T_e = 50$  ms for the cutoff time between early reflections and reverberation [20]. The late RIRs are convolved with a speech signal, transformed into the STFT domain, and the spatial coherence is estimated from auto- and cross-power spectra estimated by averaging over an interval of 10 s.

First, RIRs are generated using the image method [25], [38]. In the simulations, a uniform linear array (inter-microphone spacing  $d = 8$  cm) is placed horizontally in the center of rectangular rooms with varying dimensions and reflectivities. The image source order is chosen sufficiently high to include all reflections within 60 dB of the main peak. In order to reduce the variance of the estimate for a better visualization, the

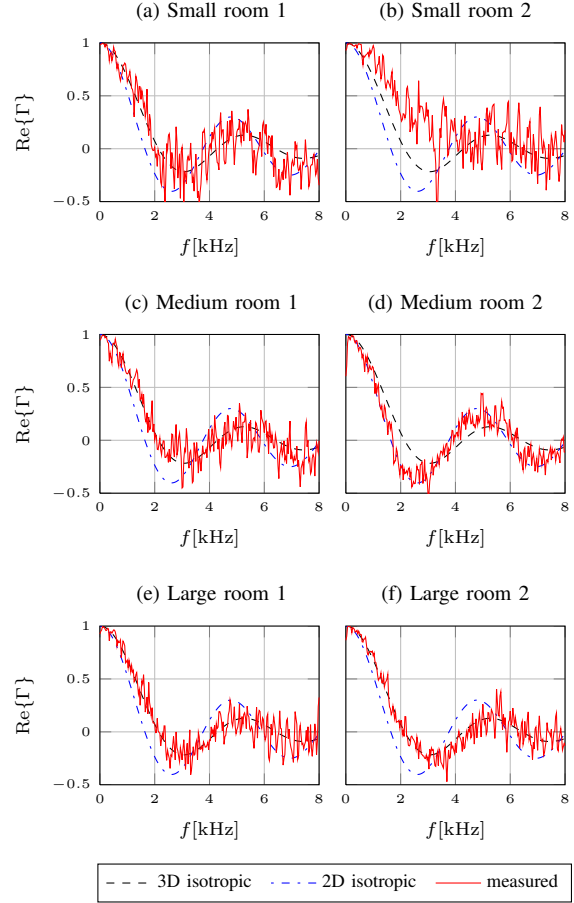


Figure 6. Spatial coherence estimated from the reverberation tail of measured RIRs from the REVERB challenge, averaged over 7 microphone pairs with spacing  $d = 8$  cm.

coherence is also spatially averaged over the estimates from 7 microphone pairs [24]. Fig. 5 shows plots of the real part of the resulting coherence, for a large room ( $15 \times 18 \times 10$  m, left) and a small room ( $4 \times 3 \times 2.5$  m, right); for both rooms, three configurations for the surface reflectivity  $\beta$  are used: equally high reflectivity for all surfaces ( $\beta = 0.9$ ), highly absorbing floor and ceiling ( $\beta_{\text{Walls}} = 0.9$ ,  $\beta_{\text{Floor, Ceil}} = 0.1$ ), and moderately absorbing walls ( $\beta_{\text{Walls}} = 0.5$ ,  $\beta_{\text{Floor, Ceil}} = 0.9$ ). The results in Fig. 5 confirm the assumptions on the coherence properties of reverberation that were made in Section III-B: for equal reflectivity of all surfaces, the coherence closely matches the coherence of the diffuse sound field. If floor and ceiling are highly absorbing, the model of a 2D isotropic sound field is appropriate. If instead the walls are more absorbing than floor and ceiling, the coherence is significantly higher than the diffuse coherence, since the dominating vertically propagating components are strongly correlated between the horizontally spaced microphones. Also, the variance of the coherence estimate is visibly lower in the larger room.

Fig. 6 shows the reverberation coherence estimates obtained from the RIRs of the REVERB challenge database, estimated in the same way as for the simulated RIRs. The coherence estimates are obtained for 7 pairs of neighboring microphones from the circular array and averaged. Most rooms match the



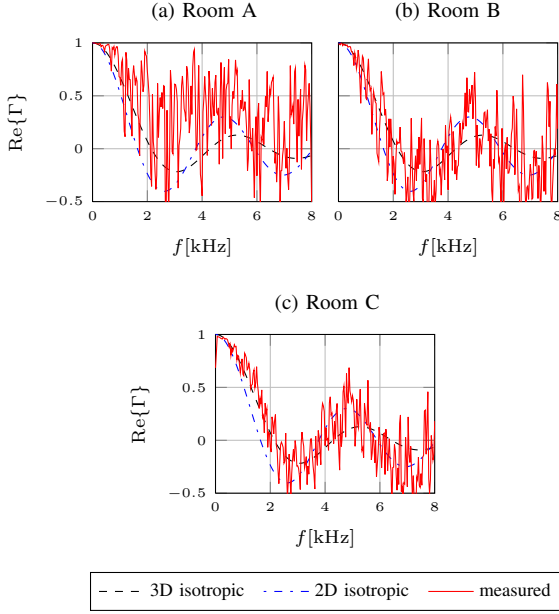


Figure 7. Spatial coherence estimated from the reverberation tail of measured RIRs in rooms A, B, C, one microphone pair with spacing  $d = 8$  cm.

diffuse model quite well, with two exceptions. In SR2, the coherence is higher than expected from the diffuse model, which can be explained by the presence of absorbing curtains on all four walls. In MR2, the coherence however almost perfectly matches the 2D isotropic model, since in this room, walls are more reflective than floor and ceiling. Also, it can again be observed that the variance of the coherence estimate is lower for rooms with a longer reverberation time.

Fig. 7 shows the results for one position in the rooms A, B and C. The coherence estimate is here computed just from one pair of microphones, therefore the variance is significantly higher. The diffuse model is a good fit for rooms B and C, where all surfaces are highly reflective. In room A, the coherence is similar to the simulated case of partially absorbing walls, which is due to the presence of partially closed curtains on the walls of the room.

Concluding the analysis of the spatial properties, it can be stated that, for microphones located in the same horizontal plane, the spatial coherence of reverberation in real rooms typically lies between the coherence of diffuse and 2D isotropic noise, with some exceptions where the coherence is increased due to dominant vertical reflections. The diffuse model is a good fit for most rooms, unless there are large differences in the reflectivity of the room surfaces. Finally, it is noteworthy that the image source model with sufficient order can reproduce the spatial characteristics of late reverberation which are observed in real rooms.

### C. CDR Estimation for Reverberant Speech

In Section II, a reverberant speech signal is modeled as consisting of a directional and a diffuse component, which are mutually uncorrelated. In practice, the reverberant sound field consists of the direct path, several spatially distinct early

reflections, and the reverberation component, all of which are not perfectly uncorrelated, due to the non-zero length of the observation window and the temporal correlation of speech signals. In the previous section, it was shown that the model of a diffuse sound field is appropriate for the reverberation component. In the following, it is investigated whether the simplified model of a mixture of uncorrelated directional and diffuse sound fields can be applied to real reverberant speech signals, i.e., whether the CDR estimate can be used as a practical measure for the time- and frequency-dependent ratio between desired and undesired signal components, as it is required for speech enhancement. We now consider the desired signal components to be the direct path plus the reflections arriving within  $T_e = 50$  ms after the direct path, and the undesired components to be the energy caused by the reverberation tail of the RIR. This is motivated by the well-known effect that early reflections are beneficial both for speech intelligibility [39] and ASR accuracy [40], and should therefore be considered part of the desired signal. In other words, the relevant SNR to be estimated for the application to signal enhancement is the early-to-late power ratio  $ELR_{50\text{ ms}}(l, f)$  (see Appendix).

To exemplarily illustrate the relationship between the (non-stationary) early-to-late power ratio and the short-time coherence estimate, the time-frequency bins of a reverberant speech signal are first classified according to the instantaneous  $ELR_{50\text{ ms}}$  into low-reverberant and highly reverberant, and the corresponding distribution of the short-time estimates of the complex coherence is visualized as a histogram. Fig. 8 shows the two-dimensional histograms of the complex coherence of bins with  $ELR > 10$  dB (left) and  $ELR < -10$  dB (right) around  $f = 1$  kHz. The coherence of the low-reverberant bins matches the coherence of a single plane wave quite well, although the signal contains contributions from early reflections in addition to the direct path. The phase has a slight spread, caused by early reflections; this has to be tolerated by the CDR estimator. The coherence of the highly reverberant bins, which should lie close to the diffuse model coherence, has a considerably higher spread and is not exactly centered around the model. This indicates that, while the simplified model seems to be reasonable, errors are non-negligible, and the differences in the realizations of the unbiased estimators, which affect only the behavior for values deviating from the ideal model, are likely to have a significant impact on estimation performance.

For the comparison of the estimation performance of the different estimators, it is convenient to transform the true and estimated CDR into the true and estimated diffuseness  $D = [CDR + 1]^{-1}$  and  $\hat{D} = [\hat{CDR} + 1]^{-1}$ , respectively, due to the diffuseness being bounded between 0 and 1, and to evaluate the mean squared error  $\overline{MSE} = \mathcal{E}\{|D - \hat{D}|^2\}$ . For this evaluation, the true CDR is again approximated by the ELR ( $CDR \approx ELR_{50\text{ ms}}$ ), and the expectation is approximated by averaging over time and frequency. The coherence models  $\tilde{\Gamma}_s$  and  $\tilde{\Gamma}_n$  for the estimators are based on the measured TDOA and the diffuse coherence assumption, respectively. Table II shows the MSE for the different estimators, averaged over all source positions in the respective room. The estimator

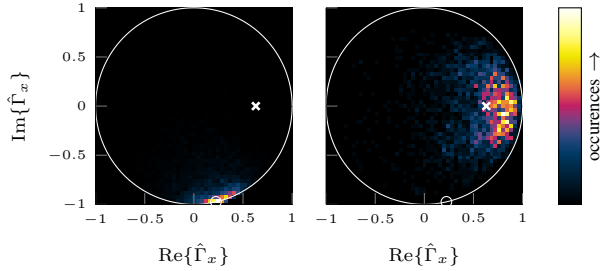


Figure 8. Histogram of complex coherence values  $\hat{\Gamma}_x$  measured from a reverberant speech signal, for time-frequency bins with  $ELR_{50\text{ ms}} > 10\text{ dB}$  (left) and  $< -10\text{ dB}$  (right). Room B,  $l = 2\text{ m}$ ,  $d = 8\text{ cm}$ ,  $\theta = 60^\circ$ ,  $f = 1\text{ kHz}$ . Theoretical signal coherence  $\Gamma_s$  computed from measured TDOA and diffuse noise coherence  $\Gamma_n$  are marked by  $\circ$  and  $\times$ , respectively.

Table II  
ESTIMATION ERROR OF DIFFERENT CDR ESTIMATORS.

CDR est.	Jeub	Thiergart 1 *	proposed 1 *	proposed 2 *	Thiergart 2	proposed 3 *	proposed 4 *
Prior inform.	DOA, $\Gamma_n$	DOA, $\Gamma_n$	DOA, $\Gamma_n$	DOA, $\Gamma_n$	$\Gamma_n$	$\Gamma_n$	DOA
Room A	0.182	0.486	0.166	0.095	0.062	<b>0.057</b>	0.243
Room B	0.146	0.301	0.140	<b>0.086</b>	0.090	0.087	0.212
Room C	0.080	0.235	0.080	<b>0.066</b>	0.103	0.104	0.159
MR1	0.131	0.373	0.114	0.069	0.059	<b>0.052</b>	0.171
MR2	0.111	0.287	0.092	<b>0.061</b>	0.073	0.066	0.159
LR1	0.119	0.313	0.109	0.068	0.067	<b>0.063</b>	0.170
LR2	0.073	0.262	0.059	<b>0.047</b>	0.071	0.069	0.134
Mean	0.120	0.322	0.109	<b>0.070</b>	0.075	0.071	0.178

\* unbiased

$\widehat{CDR}_{\text{Thiergart},1}$  has a relatively high estimation error, due to the high sensitivity of this estimator towards phase variation of the coherence. The estimator  $\widehat{CDR}_{\text{prop1}}$  shows a slightly reduced estimation error compared to the biased estimator  $\widehat{CDR}_{\text{Jeub}}$ , while the variant  $\widehat{CDR}_{\text{prop2}}$  further reduces the error. Among the DOA-independent estimators, the proposed unbiased version leads to an error reduction as well, while the noise coherence-independent variant  $\widehat{CDR}_{\text{prop4}}$  has the overall second-highest error, due to the difficulties in cases where the phase of the coherence is close to zero.

### D. Dereverberation Performance

In the following, the signal enhancement system described in Section V is evaluated for the application to dereverberation. For all of the following results, two-channel signals are processed by first applying spatial magnitude averaging as described by (28), and then applying a postfilter based on the different CDR estimators, or one of several other dereverberation methods used for comparison.

1) *Measures and Evaluation Method:* To quantify the amount of reverberation in the unprocessed and processed signals, the time- and frequency-averaged early-to-late power ratio  $ELR_{50\text{ ms}}$  is evaluated (see Appendix). The amount of signal distortion caused by the postfilter is quantified by the frequency-weighted segmental signal-to-distortion ratio (fwSegSDR), which we define as the fwSegSNR [41] computed for the postfiltered early signal component (i.e., the signal convolved with the first 50 ms of the RIR), with the unprocessed early signal component  $Y_e$  as the reference:

$$fwSegSDR = fwSegSNR(Y_e(l, f), G(l, f)Y_e(l, f)) \quad (30)$$

The overall quality of the processed signals, including both the effects of reverberation reduction and undesired speech distortion, is evaluated using the recognition rate of an automatic speech recognizer. The ASR engine PocketSphinx [42] is used with an acoustic model trained on clean speech from the GRID corpus [43], using MFCC+ $\Delta$ + $\Delta\Delta$  features. Cepstral mean normalization is used for the equalization of the effect of early reverberation [44]. For the computation of the recognition rate, only the letter and the number in the utterance are evaluated, as in the CHiME challenge [45]. Furthermore, two signal-based measures for the overall speech quality are evaluated, which were shown to be significantly correlated to the perceived amount of reverberation [46]: PESQ [47] and the frequency-weighted segmental signal-to-noise ratio (fwSegSNR) [41]. We use the wideband version of PESQ and give values in the MOS-LQO scale. For both PESQ and the fwSegSNR, the clean speech signal is used as reference.

CDR-based dereverberation is evaluated with all estimators discussed in this paper. In addition to the CDR-based methods, two heuristic coherence-based postfiltering methods are evaluated: a version of Allen's method [2], where the magnitude of the coherence is used as a spectral gain and applied to the spatially preprocessed signal, and the coherence-to-gain-mapping proposed by Westermann et al. [7], which depends on a histogram of the magnitude squared coherence. Also evaluated is the exponential decay model by Lebart et al. [48], using the true reverberation times measured from the RIRs, which in practice would have to be estimated blindly from the reverberant signals [49]. For the method of Lebart and the CDR-based methods, spectral magnitude subtraction according to (29) is applied, with  $G_{\min} = 0.1$ . The suppression parameter  $\mu$  is set to 1.3, which yields close to optimum recognition rates for all except Lebart's method (see the comment in the following section). Ideal TDOA knowledge is assumed for the CDR estimators which require a TDOA estimate  $\hat{\Delta t}$ , i.e.,  $\hat{\Delta t} = \Delta t$ . The dereverberation methods are evaluated for the rooms A, B, C, MR1/2 and LR1/2. In SR1/2, the very low amount of reverberation ( $T_{60} < 0.3\text{ s}$ ) did not lead to a significantly lower recognition rate compared to clean speech, therefore these rooms are not included in the evaluation. For each room and source position, 500 GRID utterances are convolved with the measured two-channel RIRs (in the case of the REVERB challenge RIRs, two neighboring microphones are selected from the circular array), and then processed by the dereverberation methods.

2) *Results:* Table III summarizes the resulting performance measurements, averaged over all source positions in each room. The first column shows the results for the unprocessed microphone signals. The spatial magnitude averaging leads to a small but consistent improvement in all performance measures, as seen in the second column.

Postfiltering using the CDR estimator  $\widehat{CDR}_{\text{prop2}}$  leads to the highest recognition rate among all methods across all evaluated rooms, as well as to the highest average PESQ score. Comparing the CDR-based methods, the following observations can be made: both for the DOA-dependent and DOA-independent estimators, all measures reflect the slight advantage of the respective unbiased variant ( $\widehat{CDR}_{\text{prop1}}$  and

Table III

PERFORMANCE MEASURES, AVERAGED OVER ALL SOURCE POSITIONS IN EACH ROOM. FIRST COLUMN: UNPROCESSED MICROPHONE SIGNAL, SECOND COLUMN: SPATIALLY AVERAGED MAGNITUDES WITHOUT POSTFILTERING, REMAINING COLUMNS: DIFFERENT POSTFILTERS.

Preprocessor			Squared Magnitude Averaging										
Postfilter			Lebart	Coherence-based									
				Allen	Westermann	CDR-based							
						Jeub	Thiargart 1 *	proposed 1 *	proposed 2 *	Thiargart 2	proposed 3 *	proposed 4 *	
Required	-	-	$T_{60}$	-	Coh. histog.	DOA, $\Gamma_n$	DOA, $\Gamma_n$	DOA, $\Gamma_n$	DOA, $\Gamma_n$	$\Gamma_n$	$\Gamma_n$	DOA	
Parameter	-	-	$\mu=1.3$	-	$k_p=0.30$	$\mu=1.3$	$\mu=1.3$	$\mu=1.3$	$\mu=1.3$	$\mu=1.3$	$\mu=1.3$	$\mu=1.3$	
Recognition Rate [%]	Room A	<div><div></div></div> 87.0	<div><div></div></div> 87.1	<div><div></div></div> 87.7	<div><div></div></div> 89.0	<div><div></div></div> 89.9	<div><div></div></div> 89.0	<div><div></div></div> 86.2	<div><div></div></div> 89.4	<div><div></div></div> <b>90.0</b>	<div><div></div></div> 89.8	<div><div></div></div> 89.9	<div><div></div></div> 88.2
	Room B	<div><div></div></div> 49.2	<div><div></div></div> 49.9	<div><div></div></div> 69.5	<div><div></div></div> 63.5	<div><div></div></div> 67.5	<div><div></div></div> 76.0	<div><div></div></div> 64.7	<div><div></div></div> 76.4	<div><div></div></div> <b>78.2</b>	<div><div></div></div> 72.4	<div><div></div></div> 73.0	<div><div></div></div> 67.7
	Room C	<div><div></div></div> 36.4	<div><div></div></div> 36.6	<div><div></div></div> 47.8	<div><div></div></div> 48.1	<div><div></div></div> 51.7	<div><div></div></div> 65.7	<div><div></div></div> 53.2	<div><div></div></div> 67.6	<div><div></div></div> <b>68.6</b>	<div><div></div></div> 55.8	<div><div></div></div> 56.3	<div><div></div></div> 59.5
	MR1	<div><div></div></div> 77.2	<div><div></div></div> 78.2	<div><div></div></div> 84.8	<div><div></div></div> 83.6	<div><div></div></div> 85.0	<div><div></div></div> 85.6	<div><div></div></div> 78.9	<div><div></div></div> 86.6	<div><div></div></div> <b>87.0</b>	<div><div></div></div> 86.1	<div><div></div></div> 86.3	<div><div></div></div> 84.1
	MR2	<div><div></div></div> 63.9	<div><div></div></div> 65.7	<div><div></div></div> 80.0	<div><div></div></div> 74.5	<div><div></div></div> 76.6	<div><div></div></div> 80.1	<div><div></div></div> 70.8	<div><div></div></div> 80.7	<div><div></div></div> <b>81.9</b>	<div><div></div></div> 79.8	<div><div></div></div> 80.2	<div><div></div></div> 75.9
	LR1	<div><div></div></div> 64.8	<div><div></div></div> 65.1	<div><div></div></div> 77.3	<div><div></div></div> 72.8	<div><div></div></div> 75.4	<div><div></div></div> 78.9	<div><div></div></div> 70.2	<div><div></div></div> 79.4	<div><div></div></div> <b>81.1</b>	<div><div></div></div> 77.9	<div><div></div></div> 78.8	<div><div></div></div> 75.7
	LR2	<div><div></div></div> 57.2	<div><div></div></div> 58.8	<div><div></div></div> 75.5	<div><div></div></div> 70.4	<div><div></div></div> 73.8	<div><div></div></div> 82.7	<div><div></div></div> 71.6	<div><div></div></div> 83.3	<div><div></div></div> <b>83.5</b>	<div><div></div></div> 78.6	<div><div></div></div> 78.9	<div><div></div></div> 79.4
	Mean	<div><div></div></div> 62.2	<div><div></div></div> 63.1	<div><div></div></div> 74.7	<div><div></div></div> 71.7	<div><div></div></div> 74.3	<div><div></div></div> 79.7	<div><div></div></div> 70.8	<div><div></div></div> 80.5	<div><div></div></div> <b>81.5</b>	<div><div></div></div> 77.2	<div><div></div></div> 77.6	<div><div></div></div> 75.8
	PESQ	Room A	<div><div></div></div> 1.51	<div><div></div></div> 1.53	<div><div></div></div> 1.72	<div><div></div></div> 1.58	<div><div></div></div> 1.64	<div><div></div></div> 1.67	<div><div></div></div> 1.46	<div><div></div></div> 1.67	<div><div></div></div> <b>1.76</b>	<div><div></div></div> 1.64	<div><div></div></div> 1.66
Room B		<div><div></div></div> 1.19	<div><div></div></div> 1.19	<div><div></div></div> 1.34	<div><div></div></div> 1.23	<div><div></div></div> 1.25	<div><div></div></div> 1.36	<div><div></div></div> 1.26	<div><div></div></div> 1.34	<div><div></div></div> <b>1.38</b>	<div><div></div></div> 1.27	<div><div></div></div> 1.28	<div><div></div></div> 1.29
Room C		<div><div></div></div> 1.13	<div><div></div></div> 1.13	<div><div></div></div> 1.23	<div><div></div></div> 1.14	<div><div></div></div> 1.16	<div><div></div></div> 1.31	<div><div></div></div> 1.21	<div><div></div></div> <b>1.32</b>	<div><div></div></div> 1.32	<div><div></div></div> 1.17	<div><div></div></div> 1.17	<div><div></div></div> 1.26
MR1		<div><div></div></div> 1.28	<div><div></div></div> 1.29	<div><div></div></div> <b>1.46</b>	<div><div></div></div> 1.33	<div><div></div></div> 1.41	<div><div></div></div> 1.37	<div><div></div></div> 1.26	<div><div></div></div> 1.37	<div><div></div></div> 1.45	<div><div></div></div> 1.41	<div><div></div></div> 1.43	<div><div></div></div> 1.38
MR2		<div><div></div></div> 1.30	<div><div></div></div> 1.33	<div><div></div></div> 1.56	<div><div></div></div> 1.40	<div><div></div></div> 1.48	<div><div></div></div> 1.43	<div><div></div></div> 1.28	<div><div></div></div> 1.45	<div><div></div></div> <b>1.57</b>	<div><div></div></div> 1.56	<div><div></div></div> 1.56	<div><div></div></div> 1.50
LR1		<div><div></div></div> 1.18	<div><div></div></div> 1.19	<div><div></div></div> <b>1.33</b>	<div><div></div></div> 1.21	<div><div></div></div> 1.25	<div><div></div></div> 1.24	<div><div></div></div> 1.18	<div><div></div></div> 1.22	<div><div></div></div> 1.27	<div><div></div></div> 1.24	<div><div></div></div> 1.25	<div><div></div></div> 1.25
LR2		<div><div></div></div> 1.28	<div><div></div></div> 1.31	<div><div></div></div> 1.57	<div><div></div></div> 1.37	<div><div></div></div> 1.50	<div><div></div></div> 1.54	<div><div></div></div> 1.27	<div><div></div></div> 1.57	<div><div></div></div> <b>1.61</b>	<div><div></div></div> 1.58	<div><div></div></div> 1.58	<div><div></div></div> 1.54
Mean		<div><div></div></div> 1.27	<div><div></div></div> 1.28	<div><div></div></div> 1.46	<div><div></div></div> 1.32	<div><div></div></div> 1.38	<div><div></div></div> 1.42	<div><div></div></div> 1.27	<div><div></div></div> 1.42	<div><div></div></div> <b>1.48</b>	<div><div></div></div> 1.41	<div><div></div></div> 1.42	<div><div></div></div> 1.41
fwSegSNR		Room A	<div><div></div></div> 6.15	<div><div></div></div> 6.58	<div><div></div></div> 8.34	<div><div></div></div> 7.94	<div><div></div></div> <b>8.96</b>	<div><div></div></div> 7.17	<div><div></div></div> 7.14	<div><div></div></div> 8.63	<div><div></div></div> 8.48	<div><div></div></div> 8.71	<div><div></div></div> 8.73
	Room B	<div><div></div></div> 2.07	<div><div></div></div> 2.15	<div><div></div></div> <b>6.13</b>	<div><div></div></div> 4.20	<div><div></div></div> 3.92	<div><div></div></div> 4.46	<div><div></div></div> 4.15	<div><div></div></div> 5.81	<div><div></div></div> 5.45	<div><div></div></div> 5.38	<div><div></div></div> 5.40	<div><div></div></div> 4.04
	Room C	<div><div></div></div> 1.08	<div><div></div></div> 1.31	<div><div></div></div> <b>4.58</b>	<div><div></div></div> 2.89	<div><div></div></div> 3.20	<div><div></div></div> 2.42	<div><div></div></div> 2.46	<div><div></div></div> 3.79	<div><div></div></div> 3.60	<div><div></div></div> 3.93	<div><div></div></div> 3.93	<div><div></div></div> 2.32
	MR1	<div><div></div></div> 6.97	<div><div></div></div> 7.09	<div><div></div></div> 7.94	<div><div></div></div> 7.58	<div><div></div></div> <b>8.51</b>	<div><div></div></div> 6.72	<div><div></div></div> 6.21	<div><div></div></div> 7.41	<div><div></div></div> 7.76	<div><div></div></div> 7.86	<div><div></div></div> 7.88	<div><div></div></div> 7.20
	MR2	<div><div></div></div> 5.63	<div><div></div></div> 5.84	<div><div></div></div> 7.28	<div><div></div></div> 6.81	<div><div></div></div> <b>7.64</b>	<div><div></div></div> 6.68	<div><div></div></div> 5.85	<div><div></div></div> 7.31	<div><div></div></div> 7.52	<div><div></div></div> 7.59	<div><div></div></div> 7.59	<div><div></div></div> 7.15
	LR1	<div><div></div></div> 5.65	<div><div></div></div> 5.66	<div><div></div></div> 6.75	<div><div></div></div> 6.16	<div><div></div></div> <b>7.07</b>	<div><div></div></div> 6.11	<div><div></div></div> 5.56	<div><div></div></div> 6.44	<div><div></div></div> 6.79	<div><div></div></div> 6.67	<div><div></div></div> 6.67	<div><div></div></div> 6.54
	LR2	<div><div></div></div> 5.12	<div><div></div></div> 5.55	<div><div></div></div> 7.59	<div><div></div></div> 7.24	<div><div></div></div> 8.35	<div><div></div></div> 6.98	<div><div></div></div> 6.87	<div><div></div></div> <b>8.91</b>	<div><div></div></div> 8.76	<div><div></div></div> 8.65	<div><div></div></div> 8.66	<div><div></div></div> 7.23
	Mean	<div><div></div></div> 4.67	<div><div></div></div> 4.88	<div><div></div></div> 6.94	<div><div></div></div> 6.12	<div><div></div></div> 6.81	<div><div></div></div> 5.79	<div><div></div></div> 5.46	<div><div></div></div> 6.90	<div><div></div></div> 6.91	<div><div></div></div> 6.97	<div><div></div></div> <b>6.98</b>	<div><div></div></div> 5.92
	ELR <sub>sum</sub>	Room A	<div><div></div></div> 11.21	<div><div></div></div> 11.22	<div><div></div></div> 16.77	<div><div></div></div> 11.54	<div><div></div></div> 13.80	<div><div></div></div> <b>17.33</b>	<div><div></div></div> 14.16	<div><div></div></div> 15.80	<div><div></div></div> 15.95	<div><div></div></div> 14.77	<div><div></div></div> 14.66
Room B		<div><div></div></div> 4.39	<div><div></div></div> 4.41	<div><div></div></div> <b>12.10</b>	<div><div></div></div> 4.87	<div><div></div></div> 6.41	<div><div></div></div> 11.66	<div><div></div></div> 8.17	<div><div></div></div> 10.47	<div><div></div></div> 10.43	<div><div></div></div> 8.68	<div><div></div></div> 8.54	<div><div></div></div> 8.69
Room C		<div><div></div></div> 1.03	<div><div></div></div> 0.99	<div><div></div></div> <b>8.94</b>	<div><div></div></div> 1.21	<div><div></div></div> 2.52	<div><div></div></div> 8.67	<div><div></div></div> 5.35	<div><div></div></div> 7.60	<div><div></div></div> 7.50	<div><div></div></div> 4.04	<div><div></div></div> 3.94	<div><div></div></div> 6.34
MR1		<div><div></div></div> 6.37	<div><div></div></div> 6.56	<div><div></div></div> <b>11.90</b>	<div><div></div></div> 6.71	<div><div></div></div> 8.55	<div><div></div></div> 11.55	<div><div></div></div> 8.60	<div><div></div></div> 9.99	<div><div></div></div> 9.86	<div><div></div></div> 9.42	<div><div></div></div> 9.28	<div><div></div></div> 8.32
MR2		<div><div></div></div> 8.46	<div><div></div></div> 8.81	<div><div></div></div> <b>14.87</b>	<div><div></div></div> 9.20	<div><div></div></div> 10.72	<div><div></div></div> 13.49	<div><div></div></div> 10.76	<div><div></div></div> 12.21	<div><div></div></div> 12.29	<div><div></div></div> 11.92	<div><div></div></div> 11.79	<div><div></div></div> 11.06
LR1		<div><div></div></div> 3.70	<div><div></div></div> 3.86	<div><div></div></div> <b>10.48</b>	<div><div></div></div> 3.98	<div><div></div></div> 5.62	<div><div></div></div> 8.47	<div><div></div></div> 5.73	<div><div></div></div> 6.79	<div><div></div></div> 6.94	<div><div></div></div> 6.35	<div><div></div></div> 6.24	<div><div></div></div> 6.06
LR2		<div><div></div></div> 7.37	<div><div></div></div> 7.61	<div><div></div></div> <b>15.70</b>	<div><div></div></div> 7.97	<div><div></div></div> 9.69	<div><div></div></div> 13.81	<div><div></div></div> 10.60	<div><div></div></div> 12.33	<div><div></div></div> 12.63	<div><div></div></div> 11.26	<div><div></div></div> 11.14	<div><div></div></div> 11.65
Mean		<div><div></div></div> 6.08	<div><div></div></div> 6.21	<div><div></div></div> <b>12.97</b>	<div><div></div></div> 6.50	<div><div></div></div> 8.19	<div><div></div></div> 12.14	<div><div></div></div> 9.05	<div><div></div></div> 10.74	<div><div></div></div> 10.80	<div><div></div></div> 9.49	<div><div></div></div> 9.37	<div><div></div></div> 9.55
fwSegSDR <sub>sum</sub>		Room A	<div><div></div></div> -	<div><div></div></div> -	<div><div></div></div> 12.31	<div><div></div></div> <b>27.47</b>	<div><div></div></div> 17.12	<div><div></div></div> 12.79	<div><div></div></div> 10.58	<div><div></div></div> 13.53	<div><div></div></div> 14.12	<div><div></div></div> 15.82	<div><div></div></div> 15.92
	Room B	<div><div></div></div> -	<div><div></div></div> -	<div><div></div></div> 8.25	<div><div></div></div> <b>21.77</b>	<div><div></div></div> 16.44	<div><div></div></div> 8.95	<div><div></div></div> 9.72	<div><div></div></div> 9.30	<div><div></div></div> 9.91	<div><div></div></div> 11.32	<div><div></div></div> 11.45	<div><div></div></div> 10.30
	Room C	<div><div></div></div> -	<div><div></div></div> -	<div><div></div></div> 6.60	<div><div></div></div> <b>24.05</b>	<div><div></div></div> 15.81	<div><div></div></div> 8.83	<div><div></div></div> 9.63	<div><div></div></div> 9.74	<div><div></div></div> 10.20	<div><div></div></div> 12.10	<div><div></div></div> 12.25	<div><div></div></div> 9.74
	MR1	<div><div></div></div> -	<div><div></div></div> -	<div><div></div></div> 11.04	<div><div></div></div> <b>25.55</b>	<div><div></div></div> 16.87	<div><div></div></div> 10.06	<div><div></div></div> 11.13	<div><div></div></div> 11.87	<div><div></div></div> 12.48	<div><div></div></div> 13.65	<div><div></div></div> 13.83	<div><div></div></div> 12.07
	MR2	<div><div></div></div> -	<div><div></div></div> -	<div><div></div></div> 10.00	<div><div></div></div> <b>24.14</b>	<div><div></div></div> 17.06	<div><div></div></div> 10.26	<div><div></div></div> 10.85	<div><div></div></div> 11.58	<div><div></div></div> 12.25	<div><div></div></div> 13.18	<div><div></div></div> 13.35	<div><div></div></div> 11.68
	LR1	<div><div></div></div> -	<div><div></div></div> -	<div><div></div></div> 9.10	<div><div></div></div> <b>24.58</b>	<div><div></div></div> 16.58	<div><div></div></div> 9.56	<div><div></div></div> 10.63	<div><div></div></div> 10.94	<div><div></div></div> 11.46	<div><div></div></div> 12.67	<div><div></div></div> 12.85	<div><div></div></div> 11.18
	LR2	<div><div></div></div> -	<div><div></div></div> -	<div><div></div></div> 8.31	<div><div></div></div> <b>25.58</b>	<div><div></div></div> 16.21	<div><div></div></div> 11.01	<div><div></div></div> 10.49	<div><div></div></div> 12.61	<div><div></div></div> 12.86	<div><div></div></div> 13.88	<div><div></div></div> 14.01	<div><div></div></div> 11.98
	Mean	<div><div></div></div> -	<div><div></div></div> -	<div><div></div></div> 9.37	<div><div></div></div> <b>24.73</b>	<div><div></div></div> 16.58	<div><div></div></div> 10.21	<div><div></div></div> 10.43	<div><div></div></div> 11.37	<div><div></div></div> 11.90	<div><div></div></div> 13.23	<div><div></div></div> 13.38	<div><div></div></div> 11.39

\* unbiased

$\widehat{CDR}_{\text{prop3}}$ , respectively) over the biased estimators. For the DOA-dependent estimator, the variant  $\widehat{CDR}_{\text{prop2}}$  further improves the result over the first proposed unbiased estimator, due to the different behavior of this estimator for coherence values which deviate from the ideal coherence model. The significant improvement suggests that further improvement may be possible by modeling these deviations statistically and explicitly optimizing the estimator for this model. Remarkable are the results of the DOA-independent estimators: without requiring any knowledge or estimation of source DOA or other parameters of the scenario, the CDR-based postfilter can significantly increase the overall signal quality according to all evaluated measures.

Compared to CDR-based dereverberation, the methods by Allen and Westermann yield a low ELR improvement, and at the same time a higher signal-to-distortion ratio. The overall improvement in recognition rate and PESQ is relatively low for both, while Westermann's method shows good results for

the fwSegSNR. The discrepancies between these measures can be explained by the different tradeoffs between reverberation suppression and signal distortion, which have different effects on the evaluated quality measures. Apparently, Allen's and Westermann's methods apply a lower overall amount of suppression, which benefits the fwSegSNR measure, but has a small effect on ASR recognition rate and PESQ.

It is noticeable that Lebart's method yields the highest ELR, but at the same time the worst signal-to-distortion ratio; this indicates that reverberation is overestimated, and consequently too much suppression is applied, possibly due to mismatch between the exponential decay assumption and the early part of the impulse responses [50]. Reducing the suppression gain to the optimum value  $\mu = 0.6$  to counter overestimation increases the mean recognition rate to 77.4%.

The estimator  $\widehat{CDR}_{\text{prop4}}$ , which makes no assumption on the noise coherence, yields on average comparable results to the other estimators, although it can not obtain usable CDR

estimates for some of the source positions where the TDOA is close to zero. To gain further insight into the behavior for different TDOAs, we evaluate the performance for the different source positions individually in the following. Fig. 9 shows the recognition rate for signals processed with the proposed unbiased estimators 2, 3 and 4 for the different source positions in rooms A, B and C. While dereverberation using the heuristic DOA-dependent estimator  $\widehat{CDR}_{\text{prop2}}$  yields the highest recognition rate in almost all cases, the DOA-independent estimator  $\widehat{CDR}_{\text{prop3}}$  also achieves a significant improvement over all angles. The estimator  $\widehat{CDR}_{\text{prop4}}$ , while not usable for DOA  $\theta = 0$  due to the disappearing imaginary part of the coherence, remarkably already achieves a significantly increased recognition rate for DOAs as small as  $10^\circ$ , and similar recognition rates as the DOA-independent estimator for higher DOAs. In Room A, where the mismatch between the diffuse assumption and the actual reverberation coherence is significant, the estimator slightly exceeds the performance of the (on average best) estimator  $\widehat{CDR}_{\text{prop2}}$  for some positions, indicating that in some scenarios it may be of advantage to use an estimator which does not assume an isotropic noise field.

Fig. 10 shows the time-averaged  $ELR_{50\text{ ms}}$  for different frequencies before and after processing for an exemplary scenario (room B,  $l = 2\text{ m}$ ,  $d = 8\text{ cm}$ ), where  $\widehat{CDR}_{\text{prop2}}$  was used for dereverberation. It can be seen that the dereverberation is most effective at frequencies above  $1000\text{ Hz}$ , but is already significant at frequencies as low as  $300\text{ Hz}$ .

## VII. CONCLUSION

Several well-known and some novel CDR estimation methods and their application to dereverberation have been investigated. Using simulated and measured RIRs for different environments, it has been confirmed that the commonly used model of a reverberant speech signal as a plane wave in diffuse noise is sufficiently accurate to justify the application of CDR-based signal enhancement to dereverberation. However, the known CDR estimators were found to be either biased or not robust enough for practical application to signal enhancement. It has been shown that several variants of unbiased estimators can be derived which improve robustness towards model errors, and that knowledge of either the signal DOA or the noise coherence is sufficient for estimation of the CDR. Employing the improved estimators for dereverberation has been shown to lead to improved dereverberation performance. Using the DOA-independent estimator, the proposed signal enhancement scheme constitutes a completely blind dereverberation system which requires no knowledge or estimation of the signal DOA.

### APPENDIX: DEFINITION OF THE ELR

Reverberant microphone signals  $x_i(t)$  can be written as a convolution of RIRs  $h_i(t)$  with a clean signal  $d(t)$ , i.e.,  $x_i(t) = h_i(t) * d(t)$ . The RIRs can be split at  $t = T_e$  into an early part containing direct path and early reflections, and a late part containing reverberation. To quantify the amount of reverberation in a signal, the early-to-late power ratio  $ELR_{T_e}$  can then be defined as the power ratio between the components created by convolution with the early RIR, and

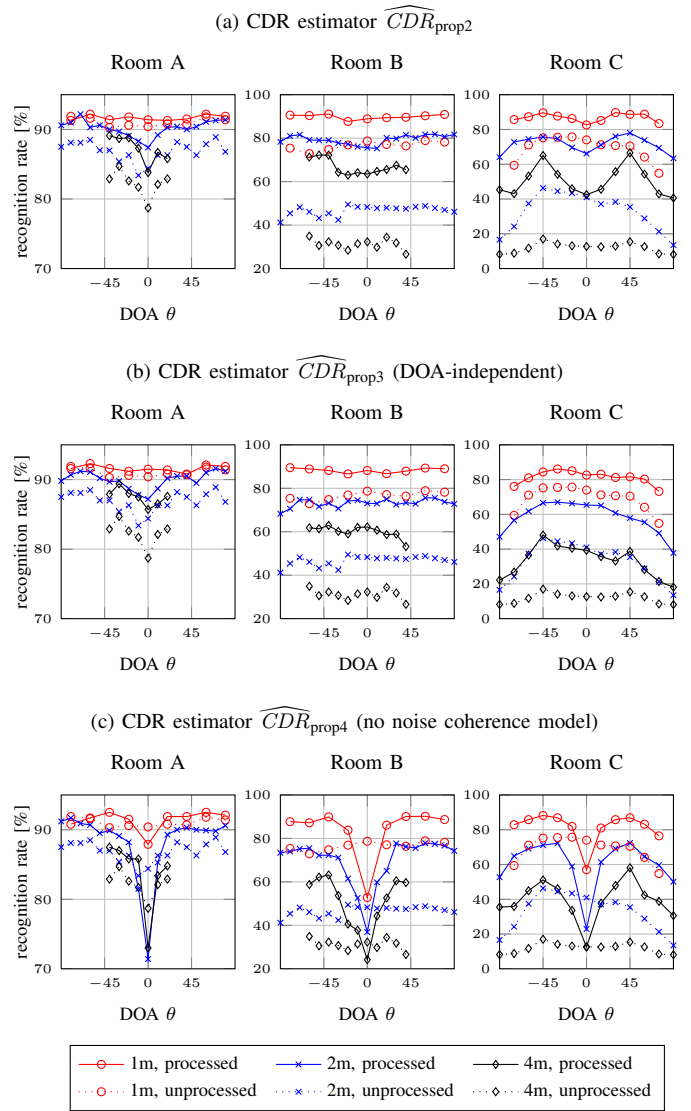


Figure 9. Average recognition rate for different rooms and source positions ( $l = 1, 2, 4\text{ m}$ ,  $\theta = -90 \dots 90^\circ$ ), for unprocessed signals and signals processed by spatial magnitude averaging combined with coherence-based postfilters based on different CDR estimators.

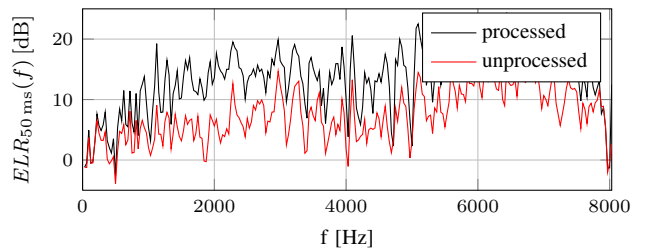


Figure 10. Time-averaged  $ELR_{50\text{ ms}}$  as function of frequency (room B,  $l = 2\text{ m}$ ,  $d = 8\text{ cm}$ ), for unprocessed reverberant signal, and signal dereverberated using the proposed unbiased estimator 2.

the reverberation components created by convolution with the late RIR, where  $T_e$  is set to an appropriate threshold, e.g.,  $T_e = 50\text{ ms}$  [20]. When  $T_e$  is set to include only the direct path in the early component, the ELR is equivalent to the DRR. For the evaluation in this paper, the  $ELR_{T_e}$  is computed for the unprocessed microphone signals, and for the signals at the



output of the signal enhancement system by processing the early and late signal components separately.

## REFERENCES

- [1] L. Danilenko, "Binaurales Hören im nichtstationären diffusen Schallfeld," *Kybernetik*, vol. 6, no. 2, pp. 50–57, Jun. 1969.
- [2] J. B. Allen, D. A. Berkley, and J. Blauert, "Multimicrophone signal-processing technique to remove room reverberation from speech signals," *J. Acoust. Soc. Am.*, vol. 62, no. 4, pp. 912–915, 1977.
- [3] P. Bloom and G. Cain, "Evaluation of two-input speech dereverberation techniques," in *Proc. ICASSP*, 1982.
- [4] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *Proc. ICASSP*, 1988.
- [5] R. Le Bouquin and G. Faucon, "Using the coherence function for noise reduction," *Communications, Speech and Vision, IEE Proceedings I*, vol. 139, no. 3, pp. 276–280, 1992.
- [6] R. Le Bouquin-Jeannes, A. Azirani, and G. Faucon, "Enhancement of speech degraded by coherent and incoherent noise using a cross-spectral estimator," *IEEE Trans. Speech and Audio Process.*, vol. 5, no. 5, pp. 484–487, Sep. 1997.
- [7] A. Westermann, J. M. Buchholz, and T. Dau, "Binaural dereverberation based on interaural coherence histograms," *J. Acoust. Soc. Am.*, vol. 133, no. 5, pp. 2767–2777, 2013.
- [8] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone Arrays*, ser. Digital Signal Process., P. M. Brandstein and D. D. Ward, Eds. Springer Berlin Heidelberg, Jan. 2001, pp. 39–60.
- [9] I. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Trans. Speech and Audio Process.*, vol. 11, no. 6, pp. 709–716, 2003.
- [10] M. Jeub, C. M. Nelke, C. Beaugeant, and P. Vary, "Blind estimation of the coherent-to-diffuse energy ratio from noisy speech signals," in *Proc. EUSIPCO*, 2011.
- [11] O. Thiergart, G. Del Galdo, and E. A. P. Habets, "Signal-to-reverberant ratio estimation based on the complex spatial coherence between omnidirectional microphones," in *Proc. ICASSP*, 2012.
- [12] —, "Diffuseness estimation with high temporal resolution via spatial coherence between virtual first-order microphones," in *Proc. WASPAA*, 2011.
- [13] —, "On the spatial coherence in mixed sound fields and its application to signal-to-diffuse ratio estimation," *J. Acoust. Soc. Am.*, vol. 132, no. 4, p. 2337, 2012.
- [14] D. P. Jarrett, O. Thiergart, E. A. P. Habets, and P. A. Naylor, "Coherence-based diffuseness estimation in the spherical harmonic domain," in *Proc. IEEE*, 2012.
- [15] A. Schwarz and W. Kellermann, "Unbiased coherent-to-diffuse ratio estimation for dereverberation," in *Proc. IWAENC*, 2014.
- [16] V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503–516, Jun. 2007.
- [17] A. Schwarz, C. Huemmer, R. Maas, and W. Kellermann, "Spatial diffuseness features for DNN-based speech recognition in noisy and reverberant environments," in *Proc. ICASSP*, 2015.
- [18] G. Del Galdo, M. Taseska, O. Thiergart, J. Ahonen, and V. Pulkki, "The diffuse sound field in energetic analysis," *J. Acoust. Soc. Am.*, vol. 131, no. 3, pp. 2141–2151, 2012.
- [19] B. F. Cron and C. H. Sherman, "Spatial-correlation functions for various noise models," *J. Acoust. Soc. Am.*, vol. 34, no. 11, pp. 1732–1736, 1962.
- [20] H. Kuttruff, *Room Acoustics*. London: Taylor & Francis, 2000.
- [21] R. Steele and L. Hanzo, *Mobile Radio Communications, 2nd Edition*. Wiley-IEEE Press, May 1999.
- [22] D. B. Kilfoyle and A. B. Baggeroer, "The state of the art in underwater acoustic telemetry," *IEEE J. Oceanic Eng.*, vol. 25, no. 1, pp. 4–27, 2000.
- [23] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. T. Jr., "Measurement of correlation coefficients in reverberant sound fields," *J. Acoust. Soc. Am.*, vol. 27, no. 6, pp. 1072–1077, 1955.
- [24] F. Jacobsen and T. Roisin, "The coherence of reverberant sound fields," *J. Acoust. Soc. Am.*, vol. 108, no. 1, pp. 204–210, 2000.
- [25] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.
- [26] G. W. Elko, E. Diethorn, and T. Gänslar, "Room impulse response variation due to thermal fluctuation and its impact on acoustic echo cancellation," in *Proc. IWAENC*, 2003.
- [27] G. W. Elko, "Superdirectional microphone arrays," in *Acoustic Signal Process. for Telecommunication*, S. L. Gay and J. Benesty, Eds. Kluwer Academic Publishers, 2000, pp. 181–237.
- [28] —, "Spatial coherence functions for differential microphones in isotropic noise fields," in *Microphone Arrays*. Springer, 2001, pp. 61–85.
- [29] E. Haensler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*. Wiley-Interscience, 2004.
- [30] M. Jeub, M. Schafer, T. Esch, and P. Vary, "Model-based dereverberation preserving binaural cues," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 18, no. 7, pp. 1732–1745, 2010.
- [31] H. Ye and R. DeGroat, "Maximum likelihood DOA estimation and asymptotic cramer-rao bounds for additive unknown colored noise," *IEEE Trans. Signal Process.*, vol. 43, no. 4, pp. 938–949, Apr. 1995.
- [32] A. Kuklasinski, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum likelihood based multi-channel isotropic reverberation reduction for hearing aids," in *Proc. EUSIPCO*, 2014.
- [33] A. Schwarz, A. Brendel, and W. Kellermann, "Coherence-based dereverberation for automatic speech recognition," in *Proc. DAGA*, 2014.
- [34] N. Ito, N. Ono, E. Vincent, and S. Sagayama, "Designing the Wiener post-filter for diffuse noise suppression using imaginary parts of inter-channel cross-spectra," in *Proc. ICASSP*, 2010.
- [35] E. A. P. Habets, "Single- and multi-microphone speech dereverberation using spectral enhancement," Ph.D. dissertation, Technische Universiteit Eindhoven, 2007.
- [36] K. Kinoshita, M. Delcroix, T. Yoshioka, T. Nakatani, A. Sehr, W. Kellermann, and R. Maas, "The REVERB challenge: A common evaluation framework for dereverberation and recognition of reverberant speech," in *Proc. WASPAA*, 2013.
- [37] M. Harteneck, S. Weiss, and R. Stewart, "Design of near perfect reconstruction oversampled filter banks for subband adaptive filters," *IEEE Trans. Circuits and Systems II: Analog and Digital Signal Process.*, vol. 46, no. 8, pp. 1081–1085, Aug. 1999.
- [38] P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *The Journal of the Acoustical Society of America*, vol. 80, no. 5, pp. 1527–1529, Nov. 1986.
- [39] J. S. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms," *J. Acoust. Soc. Am.*, vol. 113, no. 6, pp. 3233–3244, 2003.
- [40] A. Sehr, E. A. P. Habets, R. Maas, and W. Kellermann, "Towards a better understanding of the effect of reverberation on speech recognition performance," in *Proc. IWAENC*, 2010.
- [41] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech and Language Process.*, vol. 16, no. 1, pp. 229–238, Jan. 2008.
- [42] D. Huggins-Daines, M. Kumar, A. Chan, A. W. Black, M. Ravishanker, and A. I. Rudnick, "PocketSphinx: A free, real-time continuous speech recognition system for hand-held devices," in *Proc. ICASSP*, 2006.
- [43] M. Cooke, J. Barker, S. Cunningham, and X. Shao, "An audio-visual corpus for speech perception and automatic speech recognition," *J. Acoust. Soc. Am.*, vol. 120, no. 5, pp. 2421–2424, 2006.
- [44] S. Furui, "Cepstral analysis technique for automatic speaker verification," *IEEE Trans. Acoustics, Speech and Signal Process.*, vol. 29, no. 2, pp. 254–272, 1981.
- [45] H. Christensen, J. Barker, and P. Green, "The CHiME corpus: a resource and a challenge for computational hearing in multisource environments," in *Proc. Interspeech*, 2010.
- [46] S. Goetze, A. Warzybok, I. Kodrasi, J. O. Jungmann, B. Cauchi, J. RENNIES, E. A. P. Habets, A. Mertins, T. Gerkmann, S. Doclo, and B. Kollmeier, "A study on speech quality and speech intelligibility measures for quality assessment of single-channel dereverberation algorithms," in *Proc. IWAENC*, 2014.
- [47] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ) - a new method for speech quality assessment of telephone networks and codecs," in *Proc. ICASSP*, 2001.
- [48] K. Lebart, J.-M. Boucher, and P. N. Denbigh, "A new method based on spectral subtraction for speech dereverberation," *Acta Acustica united with Acustica*, vol. 87, no. 3, pp. 359–366, 2001.
- [49] C. Schuldt and P. Handel, "Decay rate estimators and their performance for blind reverberation time estimation," *IEEE/ACM Trans. Audio, Speech, and Language Process.*, vol. 22, no. 8, pp. 1274–1284, Aug. 2014.
- [50] E. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Process. Letters*, vol. 16, no. 9, pp. 770–773, Sep. 2009.